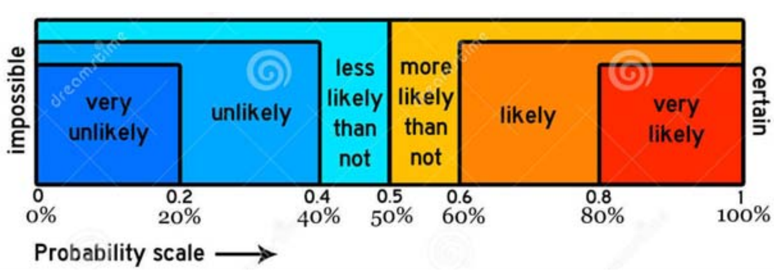
Вероватноћа и статистика

Вероватноћа – основни појмови (1)

- Вероватноћа – математичка дисциплина која се бави изучавањем случајних процеса
- Вероватноћа – квантитативна мера којом се процењује могућност/немогућност неког догађаја



Вероватноћа – основни појмови (2)

- Примена теорије вероватноће – ако не постоји могућност
  1. Детерминистичког моделирања
     - Слободан пад – детерминистички модел, може се описати једначином
     - Бацање коцкица – случајан исход који се не може се описати једначинама
  2. Мерења
     - Трајање сијалице – не могу се испитати све сијалице које нека фабрика произведе

- Примена теорије вероватноће – опис понашања неке <u>случајне променљиве</u>
  - Пробабилистичко моделирање: резултат је вероватноћа исхода

3

Случајна променљива (1)

- **Случајна променљива *X***
  - Исход неког процеса или експеримента који **<u>није унапред познат и не може се описати једначинама</u>**
    - чврстоћа бетона [GPa]
    - јачина ветра – оптерећење конструкције од ветра [kN /m^2]
    - висина падавина [mm]
    - проток [m^3/s]

- Понашање случајне променљиве се не може описати једначинама већ се анализирају подаци о тој случајној променљивој
  - Ако се експеримент понавља много пута под истим условима, појављује се одређена законитост
  - Нпр. чврстоћа бетона се одређује испитивањем бетонских коцки

4

Случајна променљива (2)

- **Случајне променљиве – класификација**

1. Дискретне (целобројне вредности): $x \in N$
   - Нпр. број возила која треба да скрену лево на раскрсници, број судара на деоници пута, број кишних дана у току године, број гласача на изборима...

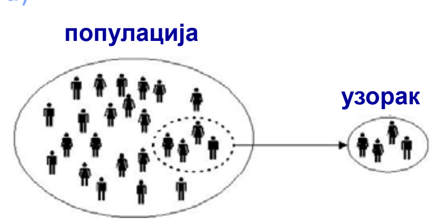
2. Континуалне (реални бројеви): $x \in R$
   - Нпр. чврстоћа бетона, оптерећење од ветра, проток, висина падавина..., општа просечна оцена или висина студената...

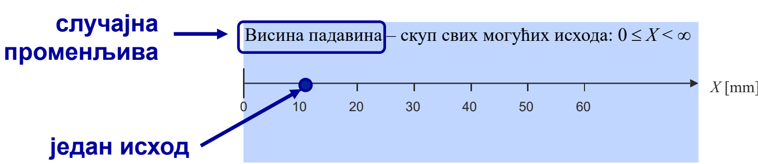
- У хидротехници се најчешће анализирају континуалне случајне променљиве (проток, водостај, висина кише, ниво подземних вода...)

Узорак и популација (1)

- Примена теорије вероватноће – заснива се на анализи узорка

- Узорак – подскуп (део) популације
  - Популација обухвата све исходе (*нпр. сви студенти ГРФа*)
  - Узорак – део популације (*нпр. К3 група*)



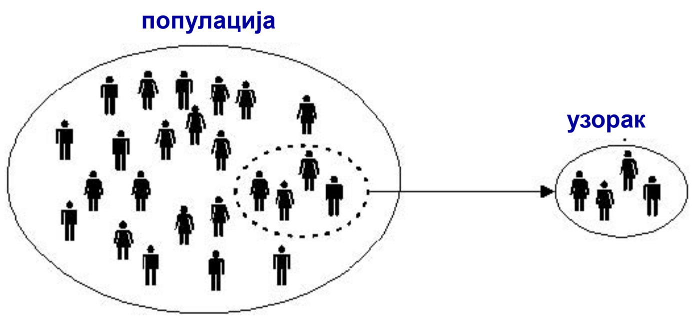


Пример: подаци о киши

Узорак: сви подаци о кишама осмотреним на некој станици

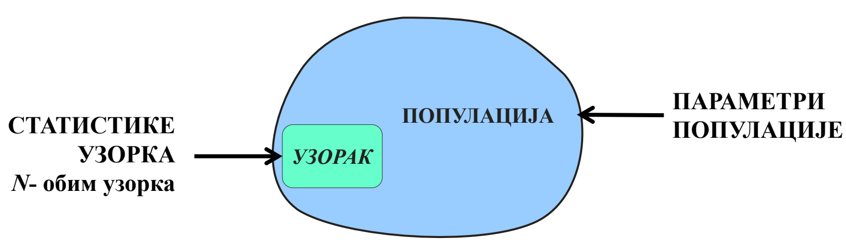
Узорак и популација (2)

- Узорак треба да буде репрезентативан
  - Хомоген: свим учесницима у анкети се поставља исто питање
  - Случајан: учесници анкете су одабрани случајно



7

Узорак и популација (3)



	СТАТИСТИКЕ УЗОРКА		ПАРАМЕТРИ ПОПУЛАЦИЈЕ
имају димензије	средња вредност:	$x_{SR} = \frac{1}{N}\sum_{i=1}^{N} x_i$	математичко очекивање: $E[x]$
	стандардна девијација:	$S_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - x_{SR})^2}$	дисперзија σ^2
немају димензије	коефицијент варијације:	$\hat{C}v = \frac{Sx}{x_{SR}}$	C_v
	коефицијент асиметрије:	$C_S = \frac{N}{(N-1)(N-2)}\sum_{i=1}^{N}\frac{(x_i - x_{SR})^3}{S_x^3}$	C_s

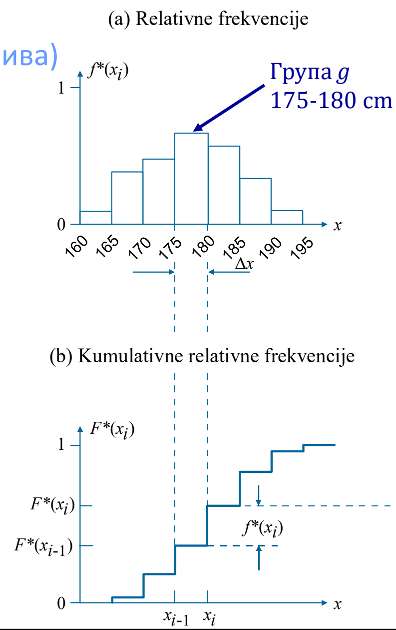
8

Релативне и кумулативне релативне фреквенције (1)

- Пример: висина студената (континуална случајна променљива)
  - У анкети је учествовало укупно N студената који су подељени у групе према висини (нпр. од 160-165 cm, 165-170 cm, итд.)
  - Релативне фркевенције: $\boxed{f^*(x_i) = \frac{N_i}{N}}$

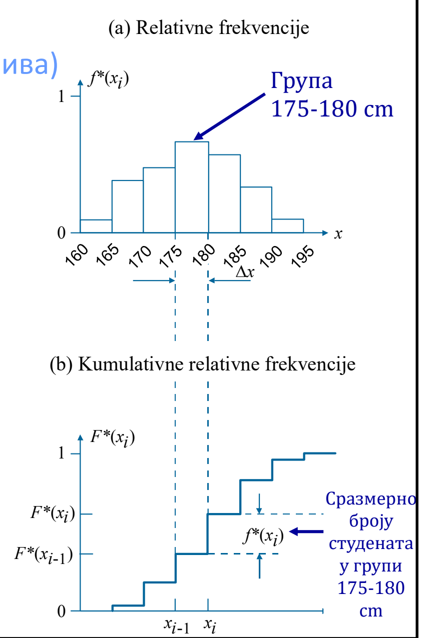
    N_i – број студената у групи i

    - Сума свих релативних фреквенција $f^*(x_i)$ је 1
    - Највећа релативна фреквенција – највише испитиваних студената је у групи g (високу су између 175 и 180 cm)
  - Кумулативне релативне фреквенције

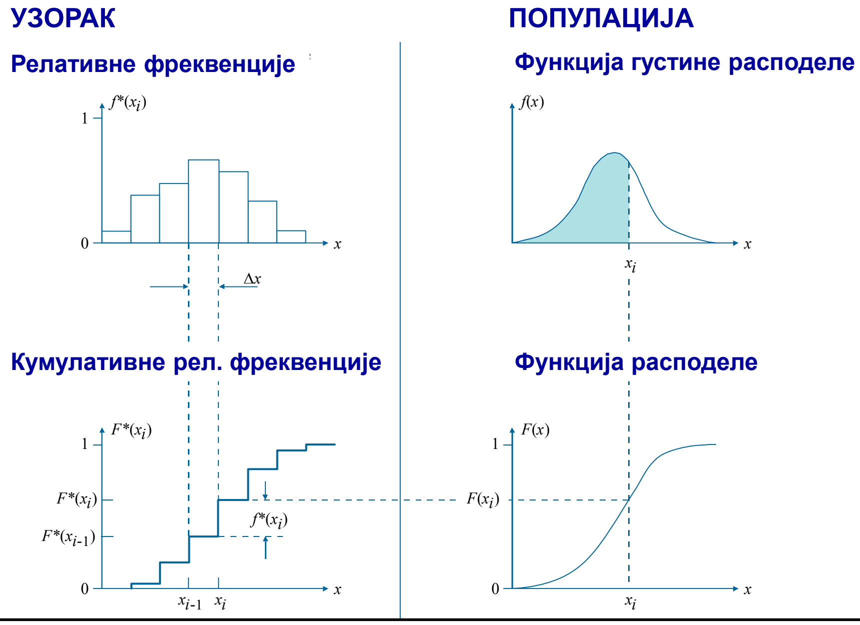


Релативне и кумулативне релативне фреквенције (2)

- Пример: висина студената (континуална случајна променљива)
  - Релативне фркевенције:
    - Сума свих релативних фреквенција је 1
  - Кумулативне релативне фреквенције: колико има студената чија је висина мања или једнака 180 cm
    - „Сумарне линије“ за релативне фреквенције
    - „Скок“ у кумулативним фреквенцијама – сразмеран броју студената у групи i
    - Највећа вредност $F^*(x_i)$ је 1



Релативне фреквенције и функције расподеле



Функције густине расподеле и функције расподеле

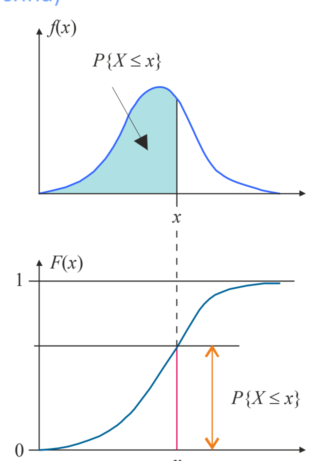
- Функције (густине) расподеле – описују понашање случајне промењљиве *x*
  - Расподеле – информације о вероватноћи исходи неког опита или процеса (нпр. вероватноћа да ниво воде у реци буде мањи или једнак коти круне насипа)
  - **Функција густине расподеле *f*(*x*)**

$$\boxed{\int_{-\infty}^{\infty} f(x)dx = 1 = 100\%} \qquad f(x) = \frac{dF(x)}{dx}$$

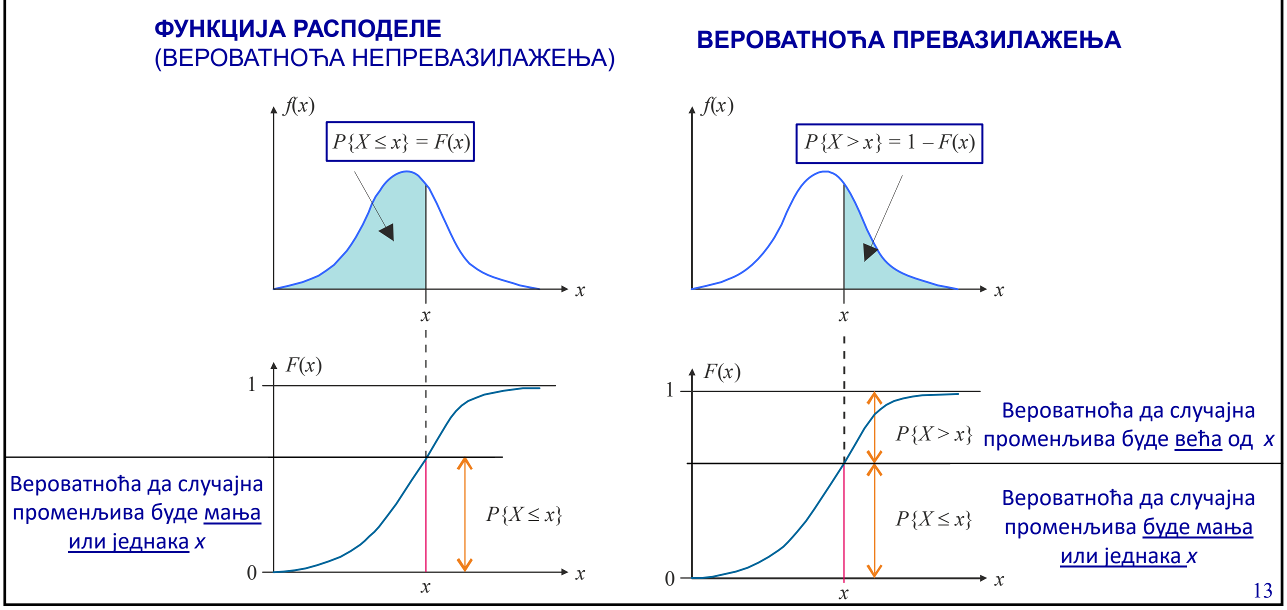
    - *f*(*x*) – непрекидна функција
    - Интеграл функције означава вероватноћу <u>непревазилажења</u>
  - **Функција расподеле *F*(*x*) (вероватноћа непревазилажења)**

$$\boxed{F(x) = P\{X \le x^*\} = \int_{-\infty}^{x^*} f(x)dx}$$

    - *F*(*x*) – монотоно неопадајућа функција



Функција расподеле и вероватноћа превазилажења

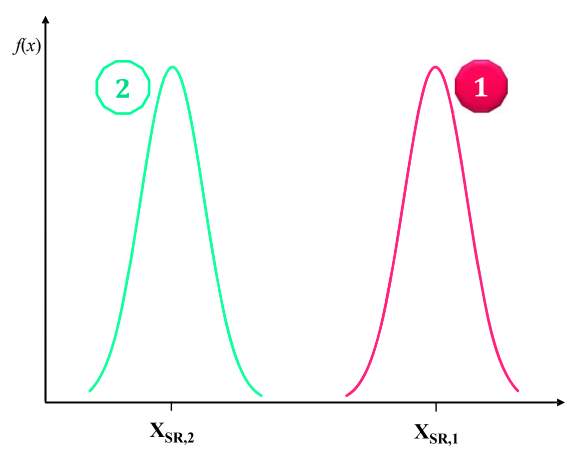


13

Особине функције густине расподеле – ср. вредност

- Особине функције расподеле – зависе од статистика узорка
- Средња вредност одређује локацију расподеле
- Пример: висина студената
  - Расподеле висине студената
  - Студенти у групи 1 су виши од студената из групе 2
    - Просечна висина студената из групе 1 је $\mathbf{X_{SR,1}}$, а суденaта из групе 2 је $\mathbf{X_{SR,2}}$

$X_{SR,1} > X_{SR,2}$



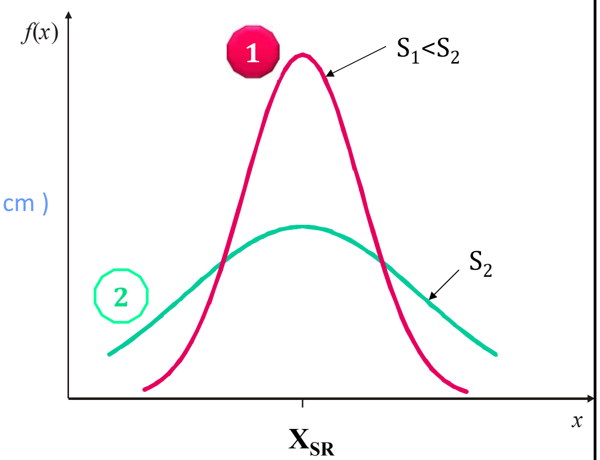
14

Особине функције густине расподеле – варијанса

- Дисперзија (варијанса) Var_X – мера одступања од средње вредности односно мера варијабилности случајне променљиве
  - Стандардна девијација S_x – корен варијансе

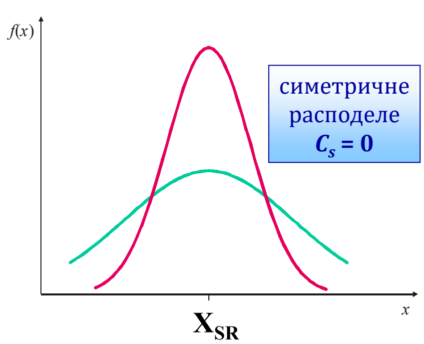
$$S_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - x_{SR})^2} \qquad Var_x = S_x^2$$

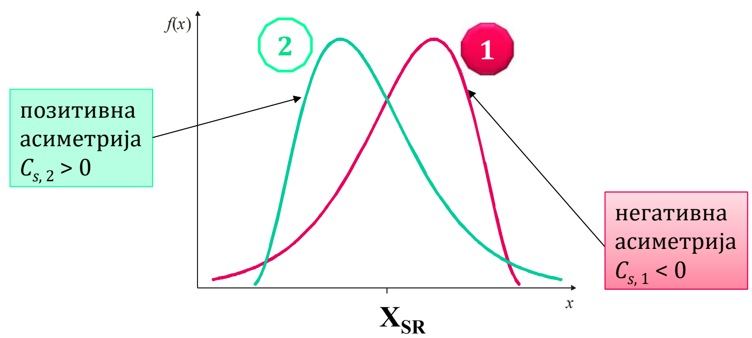
- Пример: висина студената
  - Просечна висина студената је иста у обе групе (185 cm)
  - Сви студенти из групе 1 имају приближно исту висину – мала стандардна девијација узорка (од 180 cm до 190 cm)
  - У групи 2 има много студената чија висина знатно одступа од просека – велика страндардна девијација (од 160 cm до 210 cm)



Особине функције густине расподеле – асиметрија (1)

- Асиметрија расподеле
  - коефицијент асиметрије: $C_s = \frac{N}{(N-1)(N-2)}\sum_{i=1}^{N}\frac{(x_i - x_{SR})^3}{S_x^3}$



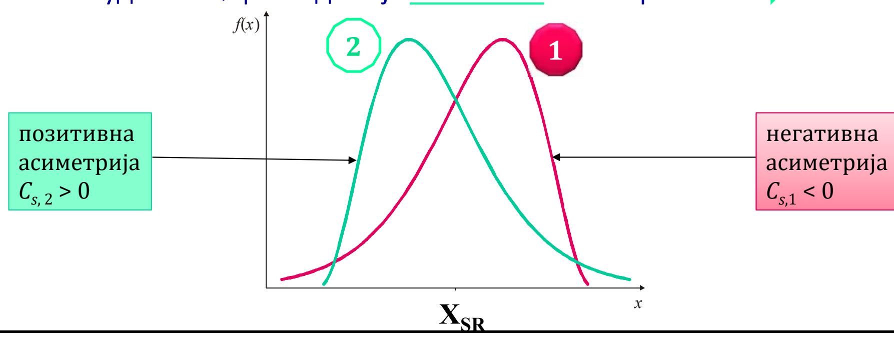


- Пример: висина студената
  - Симетричне расподеле: број студената који су виши од просека је једнак броју студената који су нижи од просека → расподела је <u>симетрична</u>

16

Особине функције густине расподеле – асиметрија (2)

- Пример: висина студената
  - Просечна висина студената у обе групе је иста
  - У групи 1 има највише студената који су виши од просека, али и неколико веома ниских студената → расподела је **негативно** асиметрична ⇐
  - У групи 2 има највише студената који су нешто нижи од просека, али и неколико веома високих студената → расподела је **позитивно** асиметрична ⇒



17

Основни појмови: $F(x)$, $P(x)$ и $R(x)$

- **Функција расподеле**: $F(x^*) = \Pr[x \le x^*]$

  → вероватноћа да случајна променљива буде мања или једнака задатој вредности x^*

- **Вероватноћа превазилажења, обезбеђеност од поплава**: $P(x^*) = \Pr[x > x^*]$

  → вероватноћа да случајна променљива буде строго већа од вредности x^*

- **Повратни период** $R(x^*)$ – просечна дужина периода између два узастопна прекорачења вредности (протока) x^*, односно просечна дужина периода током кога ће вредност x^* бити превазиђена најмање једном

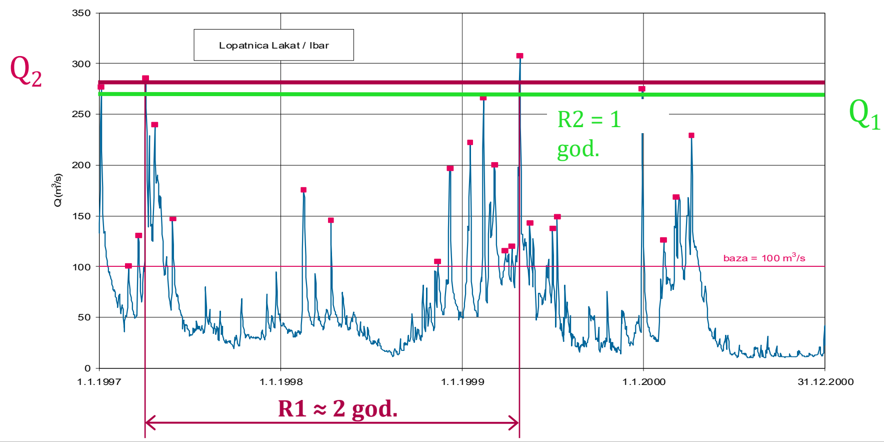
$$P(x^*) = 1/R(x^*) \quad \rightarrow \quad R(x^*) = 1/P(x^*) = 1/(1-F(x^*))$$

- Проток 100-годишњег повратног периода **НЕ ЗНАЧИ** да ће се тај проток јављати *истог дана* сваких 100 година

18

Повратни период R(x)

- o Проток Q_1 је био превазиђен 4 пута током 4 године ⟶ R(x = Q_1) ≈ 1 година
- o Проток Q_2 је био превазиђен 2 пута током 4 године ⟶ R(x = Q_2) ≈ 2 године
  - ➢ Већи протоци имају дужи повратни период $Q_2 > Q_1$



19

Теоријске расподеле

- ❑ Теоријске расподеле су дефинисане својом функцијом $F(x)$ и параметрима
- ❑ Постоји много теоријских расподела
  - o Униформна
  - o Нормална
  - o Лог-нормална
  - o Гумбелова
  - o ...

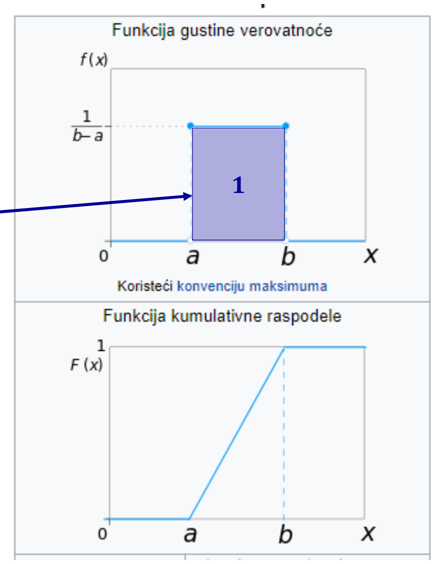
20

Униформна расподела

- Функција густине расподеле:

$$f_{\mathrm{u}}(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x < b \\ 0, & x < a,\ x \ge b \end{cases}$$

$$\int_{-\infty}^{\infty} f(x)dx = 1 = 100\%$$



- Функција расподеле:

$$F_{\mathrm{u}}(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \le x < b \\ 1, & x \ge b \end{cases}$$

Функција нормалне расподеле

- Функција нормалне расподеле:

$$\mathrm{F}_{\mathrm{norm}}(x) = \frac{1}{S_x\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left[-\frac{1}{2}\left(\frac{x-x_{\mathrm{SR}}}{S_x}\right)^2\right] dx$$

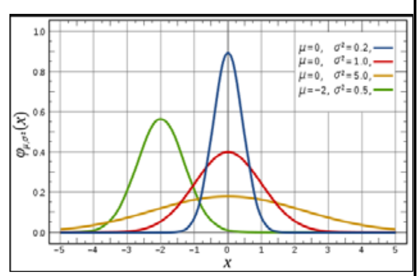
  - Не постоји аналитичко решење за интеграл функције густине нормалне расподеле

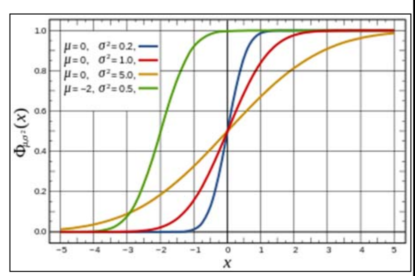
- Нормална расподела је симетрична: $C_S = 0$

- Стандардизована случајна променљива нормалне расподеле Z

$$Z = \frac{x - x_{\mathrm{SR}}}{\mathrm{S_x}}$$

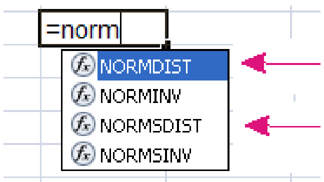
- Параметри нормалне расподеле: X_{SR} и S_x





Функција нормалне расподеле – EXCEL

- Интеграл функције нормалне расподеле се решава у EXCEL-у применом функције NORM(S)DIST
  - За дато *x* рачуна се *F*(*x*)



***F*(*x*) се рачуна на основу *x*, X_{sr} i S_x**

F*(*x*) се рачуна на основу стандардизоване случајне променљиве *z

- Функције NORM(S)INV дају вредност случајне променљиве *x* за задату функцију нормалне расподеле *F*(*x*)

Таблица нормалне расподеле

- Таблица нормалне расподеле z - F

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Друга децимала вредности *z*

Вредност F за z= 0.49

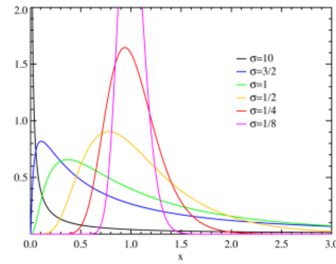
Вредности F за негативне вредности *z* се рачунају из израза: **F(-*z*) = 1 – F(*z*)**

Ако је F = 0.999, z = 3.09

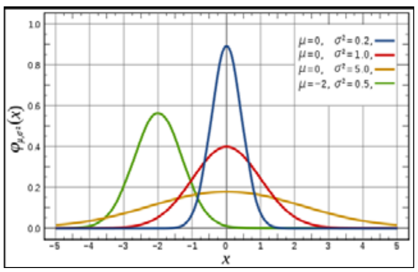
Функција лог-нормалне расподеле

- Ако логаритмоване вредности узорка прате нормалну расподелу, узорак прати лог-нормалну расподелу

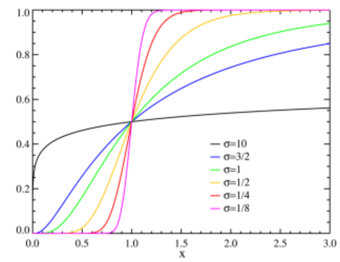
  - Лог-нормална расподела: погодна за узорке код којих је C_s логаритмованог низа је приближно једнак нули



Функција густине лог-нормалне расподеле



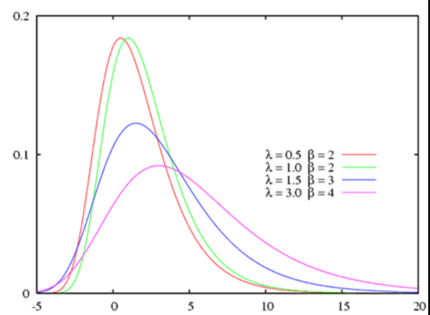
Функција густине нормалне расподеле



Функција лог-нормалне расподеле

25

Функција Гумбелове расподеле



- Гумбелова расподела:

$$F_G(x) = \exp\left(-\exp\left(-y_{G,i}\right)\right)$$

  - Функција Гумбелове расподеле – постоји аналитичко решење интеграла функције густине расподеле

- Гумбелова расподела је позитивно асиметрична $C_{s,G}$ = 1.14

- Стандардизована случајна променљива Гумбелове расподеле y_G

$$y_{G,i} = \frac{x_i - b}{a}$$

- Параметри Гумбелове расподеле: рачунају се на основу статистика узорака

$$a = 0.78\,S_x \qquad b = x_{SR} - 0.45\,S_x$$

26

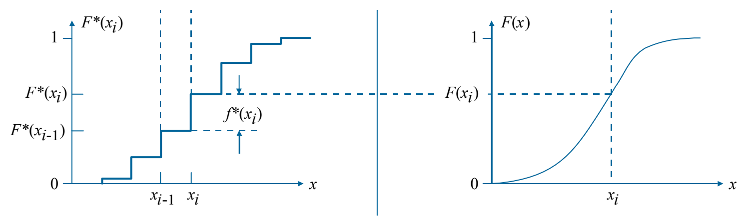
Емпиријска расподела (1)

- Кумулативна релативна фреквенција:

$$P\{X \le x_i\} = \frac{i}{N} = \frac{\text{broj podataka} \le x_i}{\text{broj podataka u nizu}}$$

  - x_i – i-ти податак у <u>низу уређеном у растући (неопадајући) редослед</u>
  - i- ранг податка у уређеном низу

Кумулативне рел. фреквенције Функција расподеле



Емпиријска расподела (2)

i	x_i	i/N
1	$x_1 = x_{min}$	1/40
2	x_2	2/40
3	x_3	3/40
4	x_4	4/40
5	x_5	5/40
...		
38	x_{38}	38/40
39	x_{39}	39/40
40	$x_{40} = x_{max}$	40/40 = 1

$$P\{X \le x_{\max}\} = \frac{N}{N} = 1$$

$$P\{X > x_{\max}\} = 1 - P\{X \le x_{\max}\} = 0$$

немогуће да X буде веће од x_{max}

?

Емпиријска расподела (3)

i	x_i	$(i-1)/N$
1	$x_1 = x_{min}$	0/40 = 0
2	x_2	1/40
3	x_3	2/40
4	x_4	3/40
5	x_5	4/40
...		
38	x_{38}	37/40
39	x_{39}	38/40
40	$x_{40} = x_{max}$	39/40

$$P\{X \le x_{\min}\} = \frac{0}{N} = 0$$

$$P\{X > x_{\min}\} = 1 - P\{X \le x_{\min}\} = 1$$

немогуће да X буде мање од x_{min}

?

29

Емпиријска расподела (4)

- “Корекција” кумулативне релативне фреквенције → емпиријска расподела (компромисне вероватноће)

Емпиријска расподела	F_{emp}
Weibull	$P\{X \le x_i\} = \frac{i}{N+1}$
Hazen	$P\{X \le x_i\} = \frac{i-0.5}{N}$
Čegodajev	$P\{X \le x_i\} = \frac{i-0.3}{N+0.4}$
Gringorton	$P\{X \le x_i\} = \frac{i-0.44}{N+0.12}$
Cuanne	$P\{X \le x_i\} = \frac{i-0.4}{N+0.2}$

30

Примена теорије вероватноће – поступак

1. Прорачун емпиријске расподеле
2. Статистике узорка и логаритмованих чланова узорка
3. Прилагођавање теоријских расподела узорку
   - Прорачун параметара расподеле на основу статистика узорка и прорачун теоријске ф-је расподеле
4. Тестирање теоријских расподела
   - Одабир (усвајање) *најпогодније* теоријске расподеле
5. Прорачун квантила датог повратног периода према *усвојеној* расподели

31

Примена теорије вероватноће – циљ

- Одабир најпогодније теоријске функције расподеле → пробабилистички модел
  - Пробабилистички модел – омогућава екстраполацију према екстремним вредностима

→ Могуће је одредити протоке великих повратних периода (нпр. 10.000 година) иако осматрања трају нпр. 25 година

32

Улазни подаци за прорачун

Hidrometrijska stanica: JASIKA
Reka: Zapadna Morava
Id. broj stanice: 47195
Površina sliva: 14721 km²
Vrsta proticaja: maksimalni godišnji (Q_{max})
Početna godina datih podataka: 1948.
Krajnja godina datih podataka: 1992.

Podaci o Q_{max} izraženi su u m³/s:

1948	899	476	327	456	804	544	658	1276	740	979	1957
1958	1328	421	464	1124	841	761	532	1870	700	1130	
	417	493	677	589	493	597	450	538	872	660	
	688	1120	905	751	477	376	593	577	1075	800	
	552	902	357	451	446	x	x	x	x	x	

Проток који је осмотрен почетне године (1948)

Формирање низа

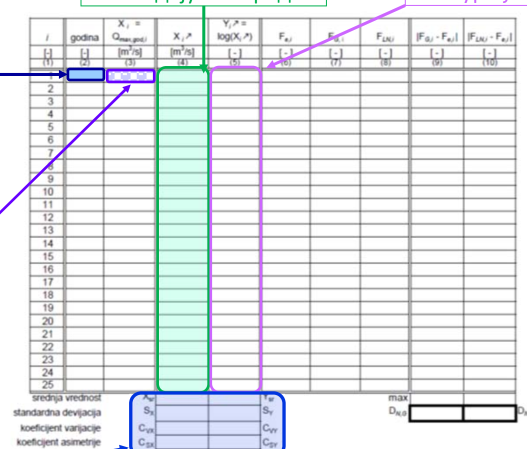
Низ протока уређен у неопадајући поредак

Декадни логаритми чланова уређеног низа (x_i↗)

Почетна година – задато у тексту задатка

Проток који одговара почетној години – дат је у табели у поставци задатка

Статистике низова



N – obim uzorka

kolona (6) kompromisna verovatnoća

$F_{eW,i} = \frac{i}{N+1}$ Weibull

$F_{eH,i} = \frac{i-0.5}{N}$ Hazen

kolona (7) Gumbelova raspodela

$F_{G,i} = e^{-e^{-y_i}}$; $y = \frac{x_i - b}{a}$

$a = 0.78\ S_X$; $b = X_{sr} - 0.45\ S_X$

kolona (8) logaritamsko-normana raspodela

$F_{LN,i} = NORMSDIST(z_i)$

$z_i = \frac{Y_i - Y_{sr}}{S_Y}$; $Y_i = \log(X_i)$

EXCEL	MatLab
X_{sr} = AVERAGE(<uzorak>)	(=mean(<uzorak>)
S_X = STDEV(<uzorak>)	(=std(<uzorak>)
C_{SX} = SKEW(<uzorak>)	(=skewness(<uzorak>))

D_N = 0.5/N + max{| $F_{T,i}$ - $F_{e,i}$ |; i=1,N}

(2) niz godina iz kojih je formiran uzorak maksimalnih godišnjih protoka
(3) slučajna promenljiva X - hronološki uređeni maksimalni godišnji protoci
(4) slučajna promenljiva X - maksimalni godišnji protoci uređeni u neopadajući niz
(5) slučajna promenljiva Y - logaritmovani maksimalni godišnji protoci uređeni u neopadajući niz (log10)
(6) kompromisna verovatnoća / empirijska funkcija raspodele za vrednost slučajne promenljive X_i↗
(7) vrednost **Gumbelove** funkcije raspodele za vrednost slučajne promenljive X_i↗
(8) vrednost **logaritamsko-normalne** funkcije raspodele za vrednost slučajne promenljive X_i↗
(9) asolutna vrednost razlike kompromisne verovatnoće i vrednosti gumbelove funkcije raspodele za vrednost slučajne promenljive X_i↗
(10) asolutna vrednost razlike kompromisne verovatnoće i vrednosti logaritamsko-normalne funkcije raspodele za vrednost slučajne promenljive X_i↗

1/2

Емпиријска расподела

Емпиријска расподела (компромисне вероватноће) према Weibull–овој **или** Hazen–овој формули

i	godina	$X_i = Q_{max,god,i}$	$X_i\nearrow$	$Y_i\nearrow = \log(X_i\nearrow)$	$F_{e,i}$	$F_{G,i}$	$F_{LN,i}$	$\lvert F_{G,i} - F_{e,i}\rvert$	$\lvert F_{LN,i} - F_{e,i}\rvert$
[-]	[-]	[m^3/s]	[m^3/s]	[-]	[-]	[-]	[-]	[-]	[-]
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
25									
srednja vrednost	X_{sr}			Y_{sr}			max		
standardna devijacija	S_X			S_Y			$D_{N,G}$		$D_{N,LN}$ LN
koeficijent varijacije	C_{VX}			C_{VY}					
koeficijent asimetrije	C_{SX}			C_{SY}					

N – obim uzorka

kolona (6) kompromisna verovatnoća

$$F_{eW,i} = \frac{i}{N+1} \quad \text{Weibull}$$

$$F_{eH,i} = \frac{i-0.5}{N} \quad \text{Hazen}$$

kolona (7) Gumbelova raspodela

$$F_{G,i} = e^{-e^{-y_i}} \; ; \; y = \frac{x_i - b}{a}$$

$$a = 0.78\ S_X \; ; b = X_{sr} - 0.45\ S_X$$

kolona (8) logaritamsko-normana raspodela

$$F_{LN,i} = NORMSDIST(z_i)$$

$$z_i = \frac{Y_i - Y_{sr}}{S_Y} \quad ; \quad Y_i = \log(X_i)$$

EXCEL	MatLab
X_{sr} = AVERAGE(<uzorak>)	(=mean(<uzorak>)
S_X = STDEV(<uzorak>)	(=std(<uzorak>)
C_{SX} = SKEW(<uzorak>)	(=skewness(<uzorak>))

$D_N = 0.5/N + \max\{\lvert F_{T,i} - F_{e,i}\rvert; i=1,N\}$

(2) niz godina iz kojih je formiran uzorak maksimalnih godišnjih protoka
(3) slučajna promenljiva X - hronološki uređeni maksimalni godišnji protoci
(4) slučajna promenljiva X - maksimalni godišnji protoci uređeni u neopadajući niz
(5) slučajna promenljiva Y - logaritmovani maksimalni godišnji protoci uređeni u neopadajući niz (log10)
(6) kompromisna verovatnoća / empirijska funkcija raspodele za vrednost slučajne promenljive $X_i\nearrow$
(7) vrednost **Gumbelove** funkcije raspodele za vrednost slučajne promenljive $X_i\nearrow$
(8) vrednost **logaritamsko-normalne** funkcije raspodele za vrednost slučajne promenljive $X_i\nearrow$
(9) asolutna vrednost razlike kompromisne verovatnoće i vrednosti gumbelove funkcije raspodele za vrednost slučajne promenljive $X_i\nearrow$
(10) asolutna vrednost razlike kompromisne verovatnoće i vrednosti logaritamsko-normalne funkcije raspodele za vrednost slučajne promenljive $X_i\nearrow$ 1/2

35

Теоријске расподеле

- Поступак прорачуна:

1. • Параметри расподеле (јединствени параметри **за узорак**)
2. • Стандардизована случајна промењива: рачуна се за сваки члан у низу
3. • Функција расподеле: рачуна се за сваки члан у низу

36

Гумбелова расподела (1)

1

- Параметри Гумбелове расподеле *a* и *b*

$$a = 0.78\, S_x \qquad b = x_{SR} - 0.45\, S_x$$

2

- Стандардизована случајна променљива y_G

$$y_{G,i} = \frac{x_i - b}{a}$$

3

- Функција расподеле

$$F_{G,i}(x) = \exp\left(-\exp\left(-y_{G,i}\right)\right)$$

37

Гумбелова расподела (2)

i **(1)**	$x_i \nearrow$ **(4)**	$y_{G,i}$	F_{GUMBEL} **(7)**
1	x_1 (Q_{min})	$y_{G,1} = \frac{x_1 - b}{a}$	$F_{G,1}(x_1) = \exp\left(-\exp\left(-y_{G,1}\right)\right)$
2	$x_2 \geq x_1$	$y_{G,2} = \frac{x_2 - b}{a}$	$F_{G,2}(x_2) = \exp\left(-\exp\left(-y_{G,2}\right)\right)$
3	$x_3 \geq x_2$	…	…

Статистике узорка

$$x_{SR} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$Sx = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - x_{SR})^2}$$

Параметри Гумбелове расподеле

$$a = 0.78\, S_x$$

$$b = x_{SR} - 0.45\, S_x$$

38

Лог-нормална расподела (1)

1
- Параметри расподеле Y_{SR} и S_Y

2
- Стандардизована случајна промењљива Z_Y

$$Z_{Y_i} = \frac{y_i - Y_{SR}}{S_Y}$$

3
- Функција расподеле F_{LN}

$$F_{LN,i}(y_i) = NORMSDIST\left(Z_{Yi}\right)$$

39

Лог-нормална расподела (2)

i **(1)**	***y*****i**↑ **(5)**	$\mathbf{Z_{Yi}}$	$\mathbf{F_{LN}}$ **(8)**
1	$y_1 = \log(x_1)$	$Z_{Y_1} = \frac{y_1 - y_{SR}}{S_Y}$	$F_{LN,1}(y_1) = NORMSDIST\left(Z_{Y_1}\right)$
2	$y_2 = \log(x_2)$	$Z_{Y_2} = \frac{y_2 - y_{SR}}{S_Y}$	$F_{LN,2}(y_2) = NORMSDIST\left(Z_{Y_2}\right)$
3	$y_3 = \log(x_3)$	...	...

- Статистике логаритмованог низа: параметри лог-нормалне расподеле

$$y_{SR} = \frac{1}{N}\sum_{i=1}^{N} y_i \qquad Sy = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - y_{SR})^2}$$

40

Квантили – протоци различитих повратних периода

- Задатак – наћи проток који одговара датом повратном периоду
- Поступак:

1
- Функција расподеле (из задатог повратног периода)
  R(*x*) = 1/(1- F(*x*)) → **F(*x*) = 1 – 1/R(*x*)**

2
- Стандардизована случајна променљива на основу срачунатог F(*x*): зависи од расподеле по којој се рачунају квантили

3
- Случајна променљива ***x*** (проток датог повратног периода – квантил): рачуна се на основу стандардизоване случајне променљиве и параметара расподеле

Прорачун квантила

- Прорачун квантила – на страни 2

Proračun vrednosti maksimalnih godišnjih protoka za različite, zadate, povratne periode.

			Gumbelova raspodela			Logaritamsko-normalna raspodela			
$R(x_i)$	$P(x_i)$	$F(x_i)$	$y_{i,G}$	$x_{i,G,rač}$	$x_{i,G,graf}$	$z_{i,N}$	y_i	$x_{i,LN,rač}$	$x_{i,LN,graf}$
[godine]	[-]	[-]	[-]	[m³/s]	[m³/s]	[-]	[log m³/s]	[m³/s]	[m³/s]
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)

$$R(x_i) = \frac{1}{P(x_i)} = \frac{1}{1-F(x_i)}$$

$$y_{i,G} = -\ln(-\ln(F_i))$$
$$x_{i,G} = a\ y_{i,G} + b$$

$$z_{i,N} = NORMSINV\ (F_i)$$
$$y_i = S_Y\ z_{i,N} + Y_{sr}$$
$$x_{i,LN} = 10^{y_i}$$

kolone:
- (1) povratni period.
- (2) obezbeđenost, verovatnoća pojave, verovatnoća prevazilaženja.
- (3) funkcija raspodele.
- (4) standardna slučajna promenljiva Gumbelove raspodele.
- (5) računska vrednost maksimalnog godišnjeg protoka određenog povratnog perioda po Gumbelovoj raspodeli.
- (6) vrednost maksimalnog godišnjeg protoka određenog povratnog perioda po Gumbelovoj raspodeli očitana sa papira verovatnoće.
- (7) standardna slučajna promenljiva normalne raspodele.
- (8) računska vrednost logaritma maksimalnog godišnjeg protoka određenog povratnog perioda po logaritamsko-normalnoj raspodeli.
- (9) računska vrednost maksimalnog godišnjeg protoka određenog povratnog perioda po logaritamsko-normalnoj raspodeli.
- (10) vrednost maksimalnog godišnjeg protoka određenog povratnog perioda po logaritamsko-normalnoj raspodeli očitana sa papira verovatnoće.

Testiranje saglasnosti teorijskih i empirijske funkcije raspodele

Kritične vrednosti testa Kolomogorova D_0 za prag značajnosti α=5% u funkciji obima uzorka *N*

α = 0.05

N	5	10	15	20	30	40	50	>50
D_0	0.56	0.41	0.34	0.29	0.24	0.21	0.19	$1.36/\sqrt{N}$

D_E = ______ Zaključak: ______________________

$D_{N,G}$ = ______ ______________________

$D_{N,LN}$ = ______ ______________________

2/2

Прорачун квантила према Гумбеловој расподели (1)

1

- Повратни период → функција расподеле (колона 3 у табели 2)

$$R(x_i) \quad \rightarrow \quad P(x_i)=\frac{1}{R(x_i)} \quad \rightarrow \quad F_G(x_i) = 1 - P(x_i)$$

2

- Стандардизована случајна промењива Гумбелове расподеле y_G (колона 4)

$$y_{G,i} = -\ln\left[-\ln\left(F_G(x_i)\right)\right]$$

3

- Стандардизована случајна промењива y_G, параметри Гумбелове расподеле *a* и *b* → случајна промењива ***x* - проток** (колона 5)

$$x_{G,i,RAC} = a\, y_{G,i} + b$$

43

Прорачун квантила према Гумбеловој расподели (2)

$R(x_i)$ (1)	$P(x_i)$ (2)	$F(x_i)$ (3)	$y_{G,I}$ (4)	$x_{G,rac}$ (5)
20	0.05	0.95	$y_{G,1} = -\ln\left(-\ln\left(F\left(x_i\right)\right)\right)$	$x_{G,i} = a\, y_{G,i} + b$
…				

Параметри Гумбелове расподеле

$$a = 0.78\, S_x$$

$$b = x_{SR} - 0.45\, S_x$$

44

Прорачун квантила према Лог-нормалној расподели (1)

1

- Повратни период → функција расподеле (колона 3 у табели 2)

$$R(x_i) \quad \rightarrow \quad P(x_i) = \frac{1}{R(x_i)} \quad \rightarrow \quad F_{\mathrm{LN}}(x_i) = 1 - P(x_i)$$

2

- Стандардизована случајна променљива F(x) → Z_{Yi} (колона 7)

$$Z_{\mathrm{Yi}} = NORMSINV\left(F_{\mathrm{LN}}(x_i)\right)$$

3

- Стандардизована случајна променљива Z_{Y}, параметри расподеле y_{SR} и S_{Y} → помоћна случајна променљива y_i (колона 8)

$$y_i = y_{\mathrm{SR}} + S_{\mathrm{Y}} \cdot Z_{\mathrm{Y},i}$$

4

- y_i је логаритам случајне променљиве x_i (колона 9)

$$x_i = 10^{y_i}$$

45

Прорачун квантила према Лог-нормалној расподели (2)

R(x_i) **(1)**	**P(x_i)** **(2)**	**F (x_i)** **(3)**	$\boldsymbol{Z_{i,\mathrm{LN}}}$ **(7)**	$\boldsymbol{Y_{i,\mathrm{LN}}}$ **(8)**	$\boldsymbol{x_{i,\mathrm{LN,rac}}}$ **(9)**
20	0.05	0.95	$Z_{i,\mathrm{LN}} = \mathrm{NORMSINV}\left(F(x_i)\right)$	$y_{i,\mathrm{LN}} = y_{\mathrm{SR}} + S_{\mathrm{Y}} \cdot Z_{i,\mathrm{LN}}$	$x_{i,\mathrm{LN}} = 10^{y_{i,\mathrm{LN}}}$
…					

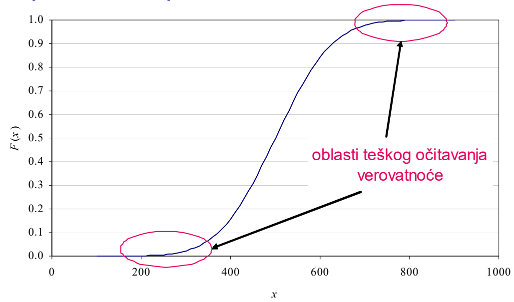
Параметри лог-нормалне расподеле:

$$y_{\mathrm{SR}} = \frac{1}{N}\sum_{i=1}^{N} y_i \qquad Sy = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - y_{SR})^2}$$

46

Графички приказ функције расподеле

- Функције расподеле су нелинеарне и квантили се тешко очитавају

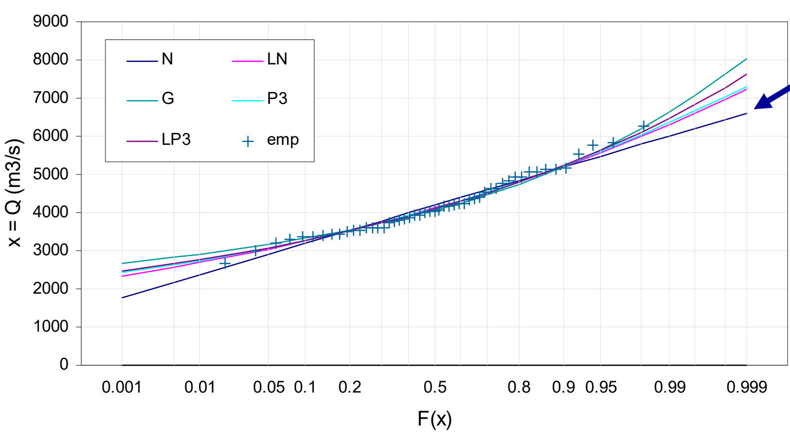


- Функције расподеле се приказују на папирима вероватноће
  - На паририма вероватноће зависност F(x) је "линеаризована" (функција расподеле је права линија и може се нацртати помоћу само две тачке)
  - Расподеле имају своје папире на којима се могу представити правим линијама (нпр. папир нормалне, лог-нормалне и Гумбелове расподеле)

47

Папир нормалне вероватноће – пример

- Различите теоријске и емпиријска расподела максималних годишњих протока осмотерених на станици Сава- Сремска Митровица
  - Папир нормалне расподеле: *симетрична* подела на апсциси

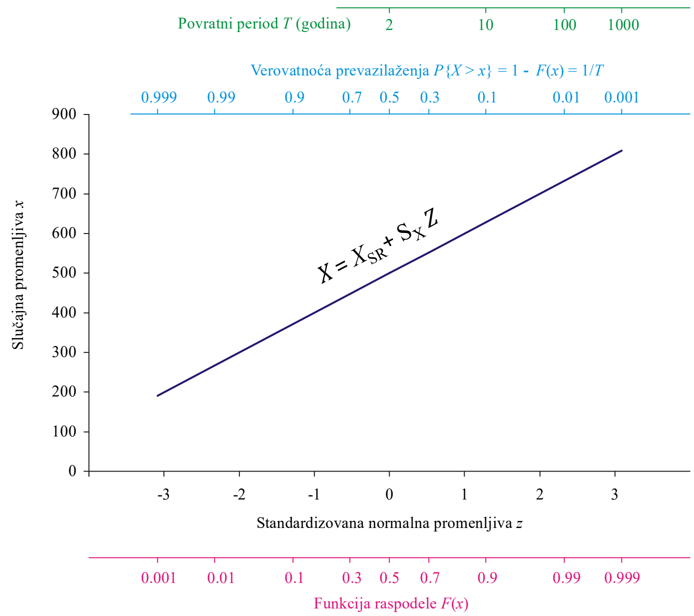


Нормална расподела је права линија на папиру нормалне вероватноће.

Остале расподеле су криве линије.

48

Папир нормалне вероватноће

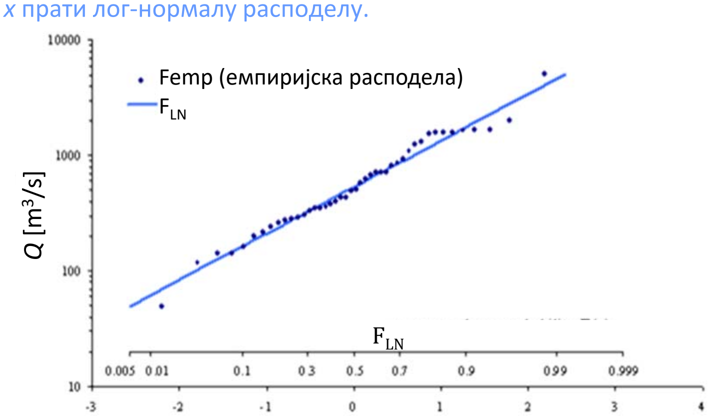


🡐 Стандардизована случајна промењива – <u>аритметичка подела</u>

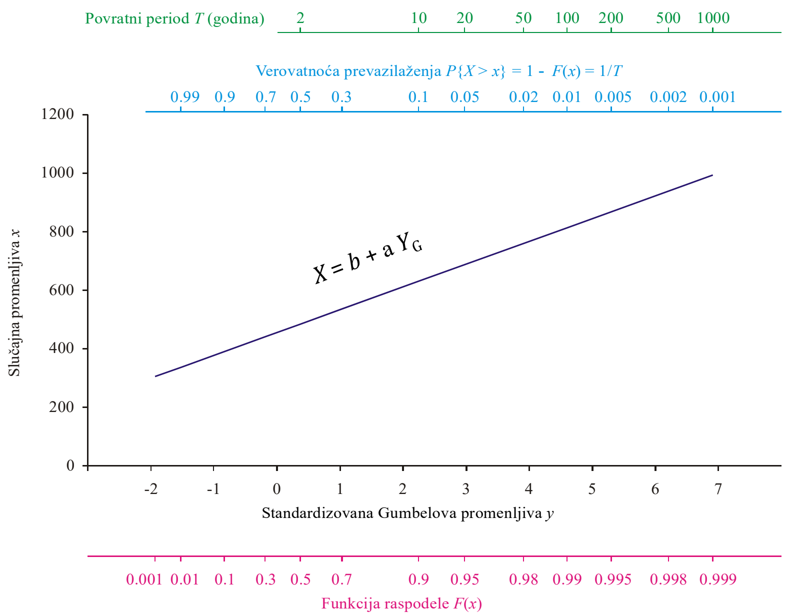
🡐 Функција расподеле – <u>нелинеарна подела</u>

Папир лог-нормалне вероватноће

- Папир лог-нормалне расподеле добија се када се вредности на ординати папира нормалне расподеле логаритмују
  - Ако логаритам случајне промењиве $y = \log x$ прати нормалну расподелу, онда случајна промењива x прати лог-нормалу расподелу.



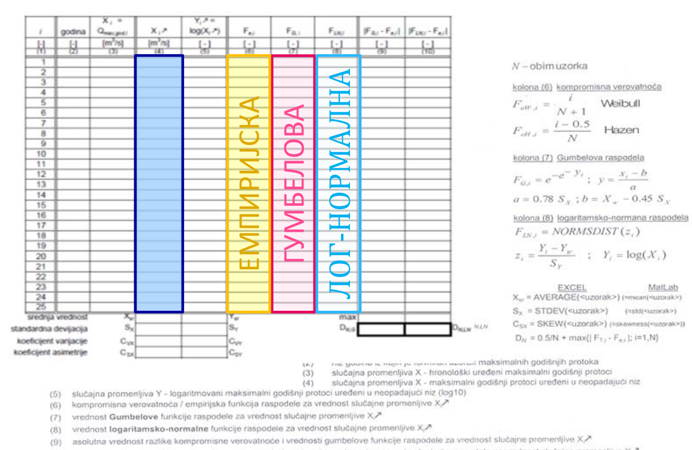
Папир Гумбелове вероватноће



51

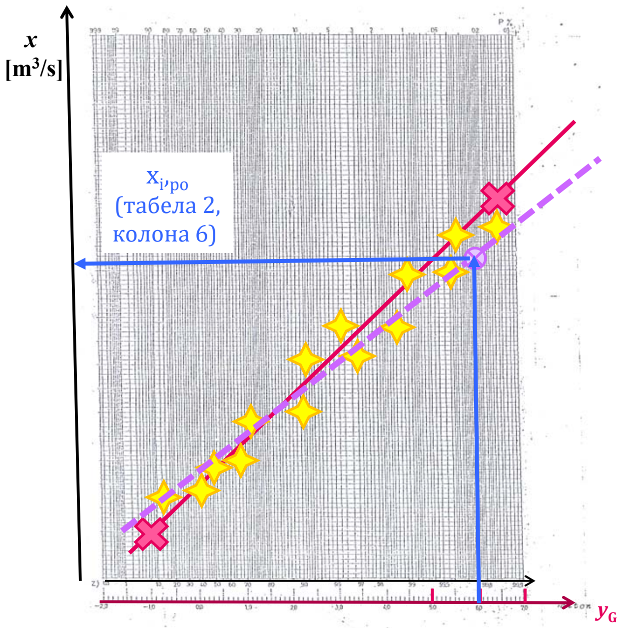
Подаци за цртање папира вероватноће

- На папирима вероватноће се цртају теоријска и емпиријска расподела
  - Теоријска расподела (Гумбелова и лог-нормална) – **права линија на одговарајућем папиру**
  - Емпиријска расподела – **низ тачака**



52

7 задатак – папир Гумбелове расподеле



Цртају се емпиријска, **теоријска рачунска** и **теоријска графичка** Гумбелова расподела

– подела на апсциси: треба да обухвати најмањи проток $X\uparrow$ и највећи квантил ($X_{g,rac}$)

– $F_{G,\ graf}$ – црта се тако да одступања од емп. расподеле буду што мања

– $x_{G,po}$ – очитава се са графички одређене теоријске Гумбелове расподеле

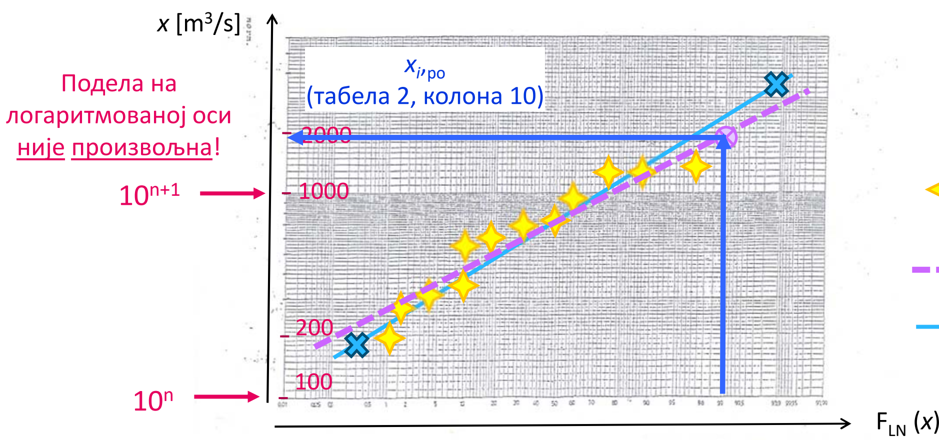
F_{EMP} (колоне 4 и 6)

$F_{G,\ graf}$ – права линија

F_{GUMBEL} (колоне 4 и 7)

Папир лог-нормалне расподеле

- Цртају се емпиријска, **теоријска рачунска** и **теоријска графичка** лог-нормална расподела
  - **Теоријска графичка** лог-нормална расподела: црта се тако да се добије најмање одступање од тачака емп. расподеле
- Са теоријске графичке расподеле очитавају се квантили за дате повратне периоде ($x_{i,LN,po}$)



F_{EMP} (колоне 4 и 6) – 25 тачака

$F_{LN,\ graf}$ – права линија

F_{LN} (колоне 4 и 8) – права линија

7 задатак – прорачун квантила

- Очитани квантили датог повратног периода се уписују у табелу 2
  - Разматрани повратни периоди су дати у поставци задатка

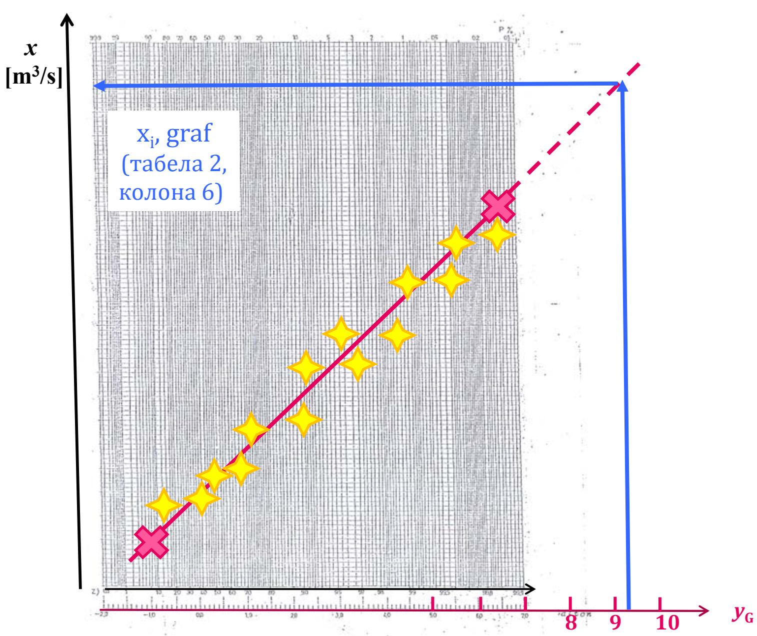
Proračun vrednosti maksimalnih godišnjih protoka za različite, zadate, povratne periode.

			Gumbelova raspodela			Logaritamsko-normalna raspodela			
$R(x_i)$	$P(x_i)$	$F(x_i)$	$y_{i,G}$	$x_{i,G,rač}$	$x_{i,G,graf}$	$z_{i,N}$	y_i	$x_{i,LN,rač}$	$x_{i,LN,graf}$
[godine]	[-]	[-]	[-]	[m^3/s]	[m^3/s]	[-]	[-]	[m^3/s]	[m^3/s]
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)

Квантили добијени очитавањем са графички одређене тероријске лог-нормалне расподеле

55

7 задатак – папир Гумбелове расподеле



Повратни период од 10.000 година
→**екстраполација до F(*x*) = 99.99%**

F_{GUMBEL} (колоне 4 и 7)

F_{EMP} (колоне 4 и 6)

$F_G(x)$ – није аритметичка подела,
не може се екстраполовати

$y_G(x)$ – стандардизована случајна променљива Гумбелове расподеле: аритметичка подела→ **екстраполација је могућа за 10000-годишњи повратни период**

56

Тестирање Колмогорова (1)

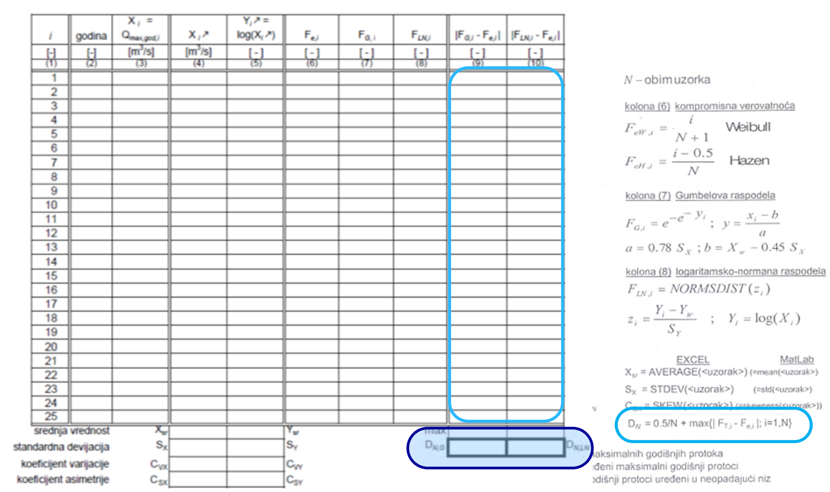
- Тестови сагласности показују да ли теоријска расподела добро апроксимира емпиријску, односно да ли су расподеле сагласне
- **Тест Колмогорова**: мери одступање теоријске и емпиријске функције расподеле
  - Конторлна статистика (максимално одступање): $D_{max} = 0.5/N + \max |F_t(x) - F_e(x)|$
  - Критеријум:
    - $D_{max} < D_0 \rightarrow$ расподеле су сагласне
    - $D_{max} > D_0 \rightarrow$ расподеле нису сагласне
  - D_0 зависи од обима узорка N и усвојеног прага зачајности α (5%)

N	α = 10%	α = 5%	α = 2%	α = 1%
20	0.265	0.294	0.329	0.352
40	0.189	0.210	0.235	0.252

  - Расподела која најбоље апроксимира емпиријску има најмању вредност D_{max}

Тест Колмогорова (2)

- Контролне статистике D_{max} за Гумбелову и лог-нормалну расподелу – одређене на основу апсолутних вредности разлика теоријских и емпиријске расподеле



Тест Колмогорова (3)

- Резултати тестирања сагласности теоријских и емпиријске расподеле

Testiranje saglasnosti teorijskih i empirijske funkcije raspodele

Kritične vrednosti testa Kolomogorova D_0 za prag značajnosti α=5% u funkciji obima uzorka N

$\alpha = 0.05$

N	5	10	15	20	30	40	50	>50
D_0	0.56	0.41	0.34	0.29	0.24	0.21	0.19	$1.36/\sqrt{N}$

Критична вредност статистике D_0 за обим узорка 25

D_0 =

$D_{N,G}$ = ________

$D_{N,LN}$ = ________

Zaključak: ____________________

Срачунате вредности конторлне статистике D_{max} за Гумбелову и лог-нормалну расподелу

1. Да ли су теоријске расподеле (Гумбелова и лог-нормална) сагласне са емпиријском?
2. Која расподела се усваја / боље слаже са емпиријском: Гумбелова **или** лог-нормална?

59

Димензионисање објеката

Proračun vrednosti maksimalnih godišnjih protoka za različite, zadate, povratne periode.

			Gumbelova raspodela			Logaritamsko-normalna raspodela			
$R(x_i)$	$P(x_i)$	$F(x_i)$	$y_{i,G}$	$x_{i,G,rač}$	$x_{i,G,graf}$	$z_{i,N}$	y_i	$x_{i,LN,rač}$	$x_{i,LN,graf}$
[godine]	[-]	[-]	[-]	[m³/s]	[m³/s]	[-]	[log m³/s]	[m³/s]	[m³/s]
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)

Квантили према различитим расподелама се разликују.
Према ком протоку треба димензионисати објекат?

- Објекти се димензионишу према квантилима који се рачунају према ***усвојеној*** расподели (резултат теста сагласности).

60